



# UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER FOR PATENTS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/600,674	07/20/2000	Antonio Pietro Sanfilippo	YAMAP0728US	9336

7590 03/18/2004

Neil A DuChez  
Renner Otto Boisselle & Sklar  
1621 Euclid Avenue 19th Floor  
Cleveland, OH 44115

EXAMINER
----------

SAIN, GAUTAM

ART UNIT	PAPER NUMBER
----------	--------------

2176

DATE MAILED: 03/18/2004

8

Please find below and/or attached an Office communication concerning this application or proceeding.

## Office Action Summary

Application No.

09/600,674

Applicant(s)

SANFILIPPO, ANTONIO PIETRO

Examiner

Gautam Sain

Art Unit

2176

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

### Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

### Status

- 1) ☒ Responsive to communication(s) filed on 20 February 2004.
- 2a) ☒ This action is FINAL. 2b) ☐ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

### Disposition of Claims

- 4) ☒ Claim(s) 1-9 is/are pending in the application.
- 4a) Of the above claim(s) \_\_\_\_\_ is/are withdrawn from consideration.
- 5) ☐ Claim(s) \_\_\_\_\_ is/are allowed.
- 6) ☒ Claim(s) 1-9 is/are rejected.
- 7) ☐ Claim(s) \_\_\_\_\_ is/are objected to.
- 8) ☐ Claim(s) \_\_\_\_\_ are subject to restriction and/or election requirement.

### Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☐ The drawing(s) filed on \_\_\_\_\_ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.  
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).  
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

### Priority under 35 U.S.C. § 119

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some \* c) ☐ None of:
- 1) ☐ Certified copies of the priority documents have been received.
  - 2) ☐ Certified copies of the priority documents have been received in Application No. \_\_\_\_\_.
  - 3) ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

\* See the attached detailed Office action for a list of the certified copies not received.

### Attachment(s)

- |  |   |
|--|---|
| 1) <input type="checkbox"/> Notice of References Cited (PTO-892)   | 4) <input type="checkbox"/> Interview Summary (PTO-413)<br>Paper No(s)/Mail Date. _____ |
| 2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948)                                   | 5) <input type="checkbox"/> Notice of Informal Patent Application (PTO-152)             |
| 3) <input type="checkbox"/> Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08)<br>Paper No(s)/Mail Date _____ | 6) <input type="checkbox"/> Other: _____  |

**DETAILED ACTION**

***Claim Rejections - 35 USC § 103***

3) The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

**3-1) Claim 1, 2, 3, 7 and 9 is rejected under 35 U.S.C. 103(a) as being unpatentable over Inaba, U.S. Patent No. 6,154,737 (issued Nov. 28, 2000), in view of Chen, European Patent Application, No. EP 0,741,364 A1 (published Jun. 6, 1996).**

In regard to independent claim 1, Inaba teaches a word frequency index for storing a frequency of occurrence of a dictionary word in the target document (Inaba, column 3, lines 64-67; compare with claim 1 "forming a structure for...strings, in which structure a string is associated with each pair of text units in which the string occurs; ...summing the number of occurrences of each other text unit in the same structure..."). Specifically, Inaba's teaching shows a structure ("frequency index") which stores the frequency of occurrences of a word (one skilled in the art would have been motivated to develop an index/structure that stores strings of words, as claimed in the invention based on the teachings of Inaba.) By comparing frequency of a word with a document, Inaba teaches a comparison of two textual objects (equivalent to a pair of text), and stores the comparison data in the frequency index. Words and document text in Inaba's

Art Unit: 2176

teachings are deemed text units. Inaba does not specifically teach a structure for the strings associated with pairs of text units. However, Chen does teach a phrase list that stores candidate phrases (Chen, column 6, lines 55-60; compare with claim 1 "...structure for each of at least some of said string "). Phrases (in Chen) are equivalent to a plurality of words (as claimed).

Additionally, Inaba teaches a frequency score calculating means for calculating the frequency of a text word occurrence in a particular document (Inabab, column 4, lines 52-56; compare with claim 1 "...form an individual score for each pair of text units"). Specifically, in Inaba, the text words and the document text form a pair of text units that are compared and the number of occurrences of the text words in the document form the score for the comparison of that particular text unit and the document text.

Additionally, Inaba teaches a word co occurrence index for storing word co occurrence (Inaba, column 4, lines 17-23, lines 30-34; compare with claim 1 "form a final score for each pair of text units to determine how many times any string is shared between each pair ..."). Specifically, a co occurrence score is the quantity of the word appearances when compared to the document text. The quantity is the final score for the comparison of that pair of text units. Inaba also teaches a degree of coincidence between the document text and the words (Inaba, lines 53-56; compare with claim language "...form a final score...").

It would have been obvious to one of ordinary skill at the time of the invention to make a text index structure, as taught in Inaba in view of Chen, that contains strings

Art Unit: 2176

and the text they are associated with, providing Inaba the benefit of indexing candidate phrases (Chen's teachings) rather than just words (Inaba's teachings) and providing the benefit of having an organized method of storing strings and the text units they associate with. It would also have been obvious to one of ordinary skill in the art at the time of the invention to interpret Inaba's system to be used for calculating individual scores of each pair of text units that contain a plurality of text, providing the benefit of comparing strings that represent a plurality of text units. Furthermore, it would have been obvious to one of ordinary skill at the time of the invention to interpret Inaba's teaching to be used for forming a final score for each string pair, providing the benefit of determining how many times any string is shared between each string pair to form a final score of co occurrence.

**In regard to dependent claim 2**, Inaba teaches a ranking means for rearranging the target document in the order of score obtained by the text unit comparison (as described in claim 1 above) (Inaba, column 4, lines 63-65; compare with claim 2 "...ranking the text units on the basis of individual scores").

**In regard to dependent claim 3**, Inaba fails to teach sentences without stop words, and stem-index records which corresponds to stem words . However, Chen teaches an automatic method for breaking document into multi-word phrases free of stop words (Chen, paragraph (57); Figure 2, items 43-58; column 1, line 45 shows 'sentences'; compare with claim language "...text units are sentences... strings are words forming said sentences, ...removing stop-words"). It would have been obvious to

Art Unit: 2176

one of ordinary skill at the time of the invention to think of Chen's teachings of multi-word phrases as equivalent to sentences (as claimed).

Additionally, Chen selects as the key phrases the candidate phrases occurring most frequently (Chen, column 2, lines 34-36; column 5, lines 45-47, figure 2, item 44; lines "...stemming each remaining word and indexing the sentences prior to carrying out said summing step..."). Additionally, Chen teaches a candidate phrase list which is a list of key phrases, excluding the stop words (Chen, column 6, lines 55-59, column 7, lines 1-4; compare with claim language "stem-index records...comprising stem words and one or more indexes corresponding to sentences ..."). A candidate phrase list (of Chen) keeps only relevant words that occur in sentences. It would have been obvious to one of ordinary skill at the time of the invention to include Chen's teachings with Inaba to develop a method of removing stop words from sentence, stemming each remaining word and thereafter creating a stem-index structures comprising stem words and the index corresponding to sentences in which stemmed words occur, providing the benefit of removing needless words and indexing only the relevant strings/sentences with their key words.

**In regard to dependent claim 7**, Inaba fails to teach the limitations of calculating a level (the highest score in relation to a threshold value) for each text unit and fails to teach a final score. However, Chen teaches a processor that compares the number of occurrences of a word within the document to a threshold and excludes those candidate terms that fall below the threshold (Chen, column 6, lines 4-10; compare with claim language "calculating a level for each text unit in addition to final score ... level indicates

Art Unit: 2176

the value of the highest of said individual scores in relation to a threshold value.”).

Specifically, to determine if a term occurs more than the threshold, the Chen processor must maintain a counter that increments whenever an occurrence is found common in comparing text units. It would have been obvious to one of ordinary skill in the art at the time of the invention, to include Chen’s teachings with Inaba, to develop a method for calculating a score, which is the number of occurrences of a word, and thereafter calculating the highest of individual scores in relation to a threshold value, providing the user the benefit of being able to generate the final scores of string comparisons and care only about those scores that are above a certain level.

**In regard to dependent claim 9**, Inaba teaches a ranking means for rearranging the target document in the order of score obtained by the text unit comparison and displays it to the user (as described in claim 1 above) (Inaba, column 4, lines 63-65; compare with claim 9 “a system for ranking text units in a text, the system comprising a data processor”). It would have been obvious to one of ordinary skill to implement the system as taught in Inaba to include a data processor programmed to perform the textual scoring/ranking operations, because data processors were very common means of performing operations such as described in the claim. The data processor provided the benefit of faster processing than doing it without a data processor.

**3-2) Claims 4 and 8 are rejected under 35 U.S.C. 103(a) as being unpatentable over Inaba, in view of Chen, as applied to claim 1 above, and further in view of Liddy, U.S. Patent No. 5,873,056 (issued Feb. 16, 1999).**

**In regard to dependent claim 4**, Inaba fails to teach words associated with subject codes. However, Liddy teaches a system that generates a subject vector representation of the text which may be a sentence, and uses subject codes that are obtained from a lexical database and assigned from the database (Liddy, Abstract section; compare with claim language "word being associated with one or more subject codes representing subjects with which said word is associated, and wherein said strings are subject codes associated with said words"). It would have been obvious to one of ordinary skill in the art at the time of the invention to include Liddy's teachings with Inaba to associate subject codes with words and strings, providing the benefit of categorizing strings and words into subject areas for more efficient searching of the strings and words.

**In regard to dependent claim 8**, Inaba does not expressly teach a storage medium. However, Liddy teaches a natural language processing system with a lexical database (Liddy, Abstract section; compare with claim language "a storage medium contain a program for controlling a programmable data process to perform a method..."). The Liddy system teaches a language processing program that controls text from documents, maintains a lexical data storage and generates subject codes.

It would have been obvious to one of ordinary skill in the art at the time of the invention to interpret Liddy's lexical database system in view of Inaba to develop a text scoring system with a database coupled with a database management system, because at the time of the invention, databases with management systems were well known in the industry as a storage medium with a programmable data process to perform a



Art Unit: 2176

method (for example, Oracle, Sybase,...). Furthermore, in Liddy, the qualification 'lexical' shows that the database is not a general purpose storage medium, rather, it is a storage medium for lexical data (similar to the claimed invention). Inherently, that shows that the database contains a programmable data processor for analyzing the data to be stored in the lexical database. This system would have provided the benefit of having a programmable data store.

**3-3) Claims 5 and 6 are rejected under 35 U.S.C. 103(a) as being unpatentable over Inaba, in view of Chen, as applied to claim 1 above, and in view of Liddy as applied to claim 4 above, and further in view of Baker, U.S. Patent No. 5,680,511 (Oct. 21, 1997).**

In regard to dependent claim 5, Inaba fails to teach a system that breaks down words into smaller components and disregards strings if the same word spelling is associated with the same subject code in a pair of string pairs. However, Baker teaches an ambiguity recognition system that recognizes a sequence of words in a document before breaking the words into smaller components and analyzing each word individually (Baker, Abstract; column 1, lines 50-57; compare with claim language "word spelling associated with each occurrence of a subject code in a text unit... occurrences of the same subject code in a pair of text units are disregarded if the same word spelling is associated with said same subject code in said pair of text units"). The examiner interprets that the claim is trying to achieve a method to reduce duplication in text units by breaking words down in smaller units in order to analyze portions (ie., analysis of letters or binary sequence of a word). Similar to this interpretation, Baker teaches a

Art Unit: 2176

word recognition system that breaks down a text unit into its components in order to compare with other text units of similar word by comparing the components.

It would have been obvious to one of ordinary skill in the art at the time of the invention to include Baker with Inaba to develop a system that breaks down words into smaller components and analyzes the components of the words (ie., analyzing spelling, binary code,...) to disregard words that have the same spelling as the subject code in the text unit pair, providing the benefit of detecting duplication and ambiguity of text words associated with subject codes.

**In regard to dependent claim 6**, Inaba fails to teach a system that does not carry out disregarding of subject codes when the codes relate to only a single word spelling in the word text. Baker teaches an ambiguity recognition system which recognizes ambiguous words that occur within a passage of words (Baker, Abstract; compare with claim language "disregarding occurrences.... single word spelling in the word text"). The examiner interprets the goal of this claim is to develop a method to reduce ambiguity when the system interprets word meanings and to not remove words that are not ambiguous. Similar to this interpretation, Baker teaches a word recognition system that reduces ambiguities amongst words in a passage of words. It would have been obvious to one of ordinary skill in the art at the time of the invention to include Baker's teaching with Inaba to develop a system that does not disregard occurrences of subject codes which relate to only a single word spelling in the word text, providing the benefit of having a data store of text units with only a single word spelling in the word text minimizing ambiguity.

***Response to Arguments***

4. Applicant' arguments filed Feb 20, 2004 have been considered, but are non-persuasive.

A. Inaba - Applicant argues "...no teaching or suggestion in Inaba et al. as to any determination of lexical cohesion among text units ... based on pair of text units". The examiner disagrees. Inaba teaches word frequency information means that is outputted to the word frequency index (Inaba, col 11, lines 42-49). Additionally, Inaba teaches word cooccurrence information in each of the documents and output it to the word cooccurrence index to make an index out ... words appearing in the same sentence ... in a cooccurrence relation to each .. in the same sentence .. extract a pair of words which are in ... relation (Inaba, col 11, lines 50-59; additionally, col 11, line 60 – col 12, line 40).

B. Inaba - Applicant argues "the present invention, on the other hand, is concerned with 'for each pair of text units (e.g., each pair of sentences).'" The examiner disagrees. Pair of text units are not limited to sentences, as pair of text units can be word, phrases, sentences (each consist one or more text units). Inaba teaches word and sentences (col 11, lines 41-60). The claim language does not exclude different documents, same document, same sentence (col 11, line 55).

C. Chen – Applicant argues The Examiner's reliance on Chen to teach a structure for string associated with pairs or text unit and argues "Chen et al. Teaches generating a list for each word of the document." The examiner argues that Inabab in view of Chen does teach a structure for strings associated with pairs of text units (Inaba suggests

Art Unit: 2176

word cooccurrence index in col 11, lines 50-55, while Chen teaches frequency of each word on the phrase list in col 5, lines 35-45). Individual and in combination, the references teach a structure (ie., list, index,...) for strings and associated with pairs of text units in which the string occurs.

D. Chen – Inaba combination – One of ordinary skill in the art would have been motivated to combine the two references because both references teach selection of key words from a document and a output of frequency analysis (Chen, col 1, line 5, lines 20-25)(Inaba, col 2, lines 60-65; summary). Inaba specifically teaches sentences (col, 11, lines 50-55) and Chen teaches phrases which can be sentences (col 5, lines 36-41).

As is argued above under section 103, it would have been obvious to one of ordinary skill in the art at the time of the invention to modify Inaba to include phrase frequency lists as taught by Chen, providing the benefit of allowing the reader to determine the content without reading the entire document and an automatic technique for generating a key word list with text understanding (Chen, col 1, lines 5-25).

E. The Examiner respectfully considered the Applicant's arguments for claims 4-6 and 8, but are non-persuasive as the claim depend on claim 1 (rejection argued above) and Liddy and Baker make up for the deficiencies (if any) in Inaba et al. and Chen et al.

### ***Conclusion***

**THIS ACTION IS MADE FINAL.** Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

Art Unit: 2176

A shortened statutory period for reply to this final action is set to expire **THREE MONTHS** from the mailing date of this action. In the event a first reply is filed within **TWO MONTHS** of the mailing date of this final action and the advisory action is not mailed until after the end of the **THREE-MONTH** shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than **SIX MONTHS** from the mailing date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Gautam Sain whose telephone number is 703-305-8777. The examiner can normally be reached on M-F 9-5 EST.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Joseph Feild can be reached on (703)305-9792. The fax phone number for the organization where this application or proceeding is assigned is (703) 872-9306.

Any inquiry of a general nature or relating to the status of this application or proceeding should be directed to the receptionist whose telephone number is (703)305-3900.

  
GS  
JOSEPH FEILD  
SUPERVISORY PATENT EXAMINER